



# The role of location in visual feature binding

Oscar Kovacs<sup>1</sup> · Irina M. Harris<sup>1</sup>

Published online: 28 January 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

Location appears to play a vital role in binding discretely processed visual features into coherent objects. Consequently, it has been proposed that objects are represented for cognition by their spatiotemporal location, with other visual features attached to this location index. On this theory, the visual features of an object are only connected via mutual location; direct binding cannot occur. Despite supporting evidence, some argue that direct binding does take over according to task demands and when representing familiar objects. The current study was developed to evaluate these claims, using a brief memory task to test for contingencies between features under different circumstances. Participants were shown a sequence of three items in different colours and locations, and then asked for the colour and/or location of one of them. The stimuli could either be abstract shapes, or familiar objects. Results indicated that location is necessary for binding regardless of the type of stimulus and task demands, supporting the proposed structure. A follow-up experiment assessed an alternate explanation for the apparent importance of location in binding; eye movements may automatically capture location information, making it impossible to ignore and suggesting a contingency that is not representative of cognitive processes. Participants were required to maintain fixation on half of the trials, with an eye tracker for confirmation. Results indicated that the importance of location in binding cannot be attributed to eye movements. Overall, the findings of this study support the claim that location is essential for visual feature binding, due to the structure of object representations.

**Keywords** Feature binding · Location · Object representation

Despite increasing evidence that visual features are processed in isolation by dedicated populations of neurons, the means by which those features are recombined into coherent objects remains unclear (Brockmole & Franconeri, 2009). Location has been repeatedly implicated in the binding of visual features (e.g., Kahneman & Treisman, 1984; Schneegans & Bays, 2017; Treisman & Gelade, 1980; Treisman & Zhang, 2006) and has consequently been assigned a central role in several theories of binding. However, conflicting evidence has persisted, suggesting that location-independent visual binding processes can take over under some circumstances (Allen, Castellà, Ueno, Hitch, & Baddeley, 2015; VanRullen, 2009).

Perhaps the best-known theory of binding is Treisman and Galade's (1980) feature integration theory (FIT). FIT contends that there are two main stages in object perception. In the first, visual features are automatically encoded in parallel, and in

the second, registered features are serially combined into objects by directing attention to a specific location. The cluster of features present at that location become bound together into a representation which can be consciously accessed. These ideas have led to the notion of object files, which occupy the middle ground between raw, unbound features and long-term memory representations (Kahneman & Treisman, 1984). Within the FIT framework, these object files are created upon attending to a specific location and are constantly updated and reviewed according to changes in perception; if a change is too large, a new object file must be created (Kahneman, Treisman, & Gibbs, 1992). These object files are thought to be integrated into a master map of spatial locations, which keeps track of all the files that are currently “open,” and through which features can be accessed.

Object indexing is a similar theory of object representation, derived from studies of infant attention. Young children lack an understanding of object permanence, and their ability to represent objects is consequently limited to those that they can presently see and manipulate (Bruce & Muhammad, 2009). For this reason, Leslie, Xu, Tremoulet, and Scholl (1998) proposed that object representation relies on mental indexes pointing to each object by its location. When an index

---

✉ Oscar Kovacs  
okov3082@uni.sydney.edu.au

<sup>1</sup> School of Psychology, University of Sydney, Sydney, NSW 2006, Australia

is created for an object, other features can be attached to that location, akin to the notion of object files. These indexes can be similarly updated and do not inherently carry any information about an object. An important claim of both theories is that each of the features of an object are bound to its location, but there is no direct connection between them (see also Schneegans & Bays, 2017).

A number of studies have provided evidence consistent with this theoretical position. For instance, Golomb, Kupitz, and Thiemann (2014) asked participants to judge whether two sequentially presented objects were the same, and found that although location was task irrelevant, participants were significantly more likely to report that identity was the same when the objects shared their location. This relationship did not extend the other way, however; shared identity did not influence judgements of location. A similar bias was found in a study by Pertzov and Husain (2014) that used bars presented in different orientations and different spatial locations. They found that subjects were more likely to subsequently misreport the orientation of a bar that shared the same spatial location as the probed target, but not that of a bar that shared the same colour as the target but which was in a different spatial location, thus indicating that spatial location supports correct feature binding. The authors suggested that their results reflect an important role for location during postperceptual processing, rather than solely during perception. There does not appear to be an equivalent dependency between other visual features; Bays, Wu, and Husain (2011) reported that judgements of the colour and orientation varied completely independently when objects were probed by location. Furthering this, Rangelov and Zeki (2014) compared the fit of colour and orientation reports to three models of binding: a strong account in which accurate judgements of both features should covary completely, an asymmetric account where the easier feature can be reported alone, and a nonbinding account in which accuracy of reports of each feature would be independent. They found that reports of colour and orientation were largely independent, and their data were best explained by the nonbinding account.

A study by Rajsic and Wilson (2014), comparing working memory access for colour and location, also suggested that location information is stored differently to other visual features. Participants viewed a set of coloured rings in different locations and could be prompted by colour or by location for the other feature of one of the rings. Rajsic and Wilson found that errors when reporting each feature were of different kinds: Mistakes reporting colour tended to be random guesses, whereas mistakes reporting location were typically swaps for the location of distractors. This pattern of reports suggests that locations are encoded in a different manner to colour: Participants may provide swaps for locations because they are only reporting indexes that were created, but for colour they would have no way of accessing unbound features and

would consequently have to guess. These findings also fit well with evidence from another recent study (Chen & Wyble, 2015), which indicated that even when instructed to ignore location, or when other features identified a target, participants were capable of reporting location on surprise questions with far greater accuracy than other features, indicating that location was automatically encoded.

Although these findings are well explained by location indexing, some research has challenged the special role of location in feature binding. Allen et al. (2015) used a visual suffix interference paradigm to test whether direct binding between features was possible. Participants were asked to remember four coloured shapes, which were followed by two competing objects (the suffixes), and then were prompted for a target from the original set on the basis of colour, shape, or location. Allen et al. reasoned that if location were automatically encoded, the suffixes would impair recall more when they appeared in the same location as the target, relative to when they shared other features with the target. However, they found this differential impairment only when the target was indicated by a location prompt; suffixes sharing their location with the target did not produce as large a deficit when the target was prompted by colour or shape instead. Because colocation of the distractors did not always differentially impair responding, Allen et al. concluded that location does not always play a special role in binding processes.

A further challenge to location-dependent accounts comes from successful binding between the features of superimposed items. Studies using random-dot kinematograms (RDK) have demonstrated that participants are able to report the direction or speed of motion, given the colour of a target (or vice versa) among overlapping populations of dots (Rodríguez, Valdes-Sosa, & Freiwald, 2002; Valdes-Sosa, Cobo, & Pinilla, 1998). In these studies, the movement within each coloured population of dots is only partially coherent, such that participants cannot not just attend to a single item or spatial location and reliably respond correctly, suggesting that binding can be successful in the absence of distinct location information. However, location-based binding strategies may still be employed in these tasks. Although the motion is not fully coherent across each population of dots, participants may be focusing on a portion of coherently moving dots restricted to a smaller area and use that location for subsequent binding purposes. Alternatively, even if participants are attending to the two dot populations in their entirety, it is not clear that they are necessarily ascribing them the same location because of the overlap. In reality, when two items are presented as overlapping, they are often perceived to be in different depth planes (this can be due to a variety of cues; see Rogers, 2009; Troscianko, Montagnon, Clerc, Malbert, & Chanteau, 1991), and these could be used to index location. Finally, a number of studies have shown that when multifeature objects overlap in the same spatial location, the successful binding of their

features is heavily dependent on the temporal parameters of the presentation (Arnold, 2005; Clifford, Holcombe, & Pearson, 2004; Holcombe, 2009; Holcombe & Clifford, 2012). Binding in this case is only possible when presentation rates are slow enough to provide additional binding cues.

## The role of object familiarity in binding

A separate aspect of location-based accounts of binding that has not been satisfactorily addressed is whether they extend to familiar objects. Most of the results that implicate location in binding come from experiments using very simple stimuli such as coloured shapes (Golomb et al., 2014; VanRullen, 2009), which bear little resemblance to real-world objects. Stimuli like this are often selected for experimental work because they have very few relevant dimensions or preexisting associations which could bias results, but they may consequently be subject to different binding processes. Memorising arbitrary pairings of colour and shape might be more likely to engage location in binding because there are so few other dimensions that features can be linked along. When dealing with richer, real-world stimuli, location might be less integral to binding, and processing may be less serial than suggested by theories such as FIT. Indeed, there is evidence that the categorisation of natural scenes occurs at speeds that are too high to be the result of serial biological processes (Kirchner & Thorpe, 2006). VanRullen (2009) consequently proposed two forms of binding: a slow, on-demand process which could be of the serial kind described by FIT and location indexing, and a second, rapid, hard-wired network developed to handle commonly encountered conjunctions without requiring substantial attentional processes. This two-process account of binding is supported by findings that transcranial magnetic stimulation (TMS) of the parietal lobe impairs search speed for novel, but not remembered, conjunctions (Walsh, Ashbridge, & Cowey, 1998). Novel conjunctions would need to be linked to locations, relying on the parietal lobe, whereas previously learned conjunctions would not.

It would make sense for hardwired conjunctions to not depend on location in the same manner as the serial binding system. Frequently encountered real-world items such as stop signs or lemons have long-term representations from which their colour can be extracted without reliance on location and spatial attention (Hommel & Colzato, 2009). This could explain why people process natural scenes extremely quickly—too quickly to be the result of binding processes (Kirchner & Thorpe, 2006)—because they are only engaging in feature perception and can fill in the other features based on long-term representations. However, familiar items with robust long-term representations may also show a reduced reliance on location even if they lack an overriding colour association, because features can be linked to that representation instead.

Essentially, when presented with simplistic line or shape stimuli in experimental work, location may be the best option for participants to link other features to; whereas for a familiar item with richer associations, the long-term representation could be used as the object index instead. This potentially reduced reliance on location for familiar stimuli without strong colour associations is, therefore, worthy of further investigation.

## The role of eye movements in indexing location

Several studies have suggested a possible functional link between eye movements and location in object representation. Patients with ophthalmoparesis (eye paralysis) appear to be significantly impaired in object processing and visual construction beyond what can be explained by perceptual difficulties (Bosbach, Kornblum, Schröder, & Wagner, 2003). Similarly, patients who suffer from Balint's syndrome as a result of bilateral lesions to the parietal lobes present with severe spatial difficulties and impairments in feature binding, as well as with oculomotor apraxia (an inability to direct their eye movements to visual targets; Coslett & Lie, 2008; Friedman-Hill, Robertson, & Treisman, 1995; Robertson & Treisman, 2006). Also, tellingly, when healthy participants are asked to report features of a previously studied item, their eyes return to the now empty location in which the target was presented, even when location is not task relevant (Meegan & Honsberger, 2005), and during recall for spatial relations bound over time, participants' gazes predictably work through sequences of objects in the order in which they were presented (Ryan & Villate, 2009).

Thus, information automatically derived from eye movements could be responsible for the apparently consistent implication of location in binding, despite claims that location is not actually necessary for feature binding (Allen et al., 2015). There is a dedicated population of neurons in the posterior parietal cortex of monkeys which produces predictable, monotonic responses to the orbital position of the eye (Zipser & Andersen, 1988), resulting in distinct firing patterns for each direction of gaze. This means that whenever the eyes are oriented towards an object, a specific signal is being produced which may then be attached to that object's representation. In other words, looking at a stimulus to encode its other features might mean that location effectively comes for free, resulting in the automatic registration of location described by Chen and Wyble (2015). If this is the case, the persistent encoding of location during binding may be a side effect of eye movement, rather than reflecting any necessary aspect of cognition.

## The current study

The aim of the present study was to evaluate location-based accounts of binding and the role of eye movements in this process. We used a short-term visual memory task, in which participants viewed sequences of three stimuli in random colours and locations. They were then prompted to report the colour and/or location of one of the items, so that we could assess any contingency between binding of location and other visual features.

To investigate the possibility that familiar objects are handled by a different binding system which is not location dependent, Experiment 1 used both photorealistic everyday objects and abstract, unfamiliar shapes as stimuli. If location is only used as an index when there is not a more developed associative framework already, any contingency between colour and location that exists for abstract shape stimuli would not be expected for realistic object stimuli.

In a second experiment, we tested whether feedback from eye movements plays a functional role in binding processes. Participants were required to maintain fixation during half of the experimental trials, with an eye tracker used to confirm that eye movements did not occur, and were free to move their eyes in the other half. A different pattern of results between restricted and free eye-movement conditions would indicate that some sort of feedback from eye movements is used in binding. If location is not actually essential for binding and simply appears important because it is difficult to ignore, we would expect a reduction in location accuracy relative to colour in the restricted eye-movement condition.

## Experiment 1

Experiment 1 was designed to test whether direct binding between visual features is possible, as contended by Allen et al. (2015), or if location is necessary for binding processes, using both familiar objects and unfamiliar shapes as stimuli. Participants were shown a sequence of three items in different colours and locations, and were then prompted by shape and asked to report the colour and/or location of the target; one block of trials required report of both colour and location,

while two other blocks required report of one of these features only. Presentation was sequential rather than simultaneous to prevent any possible advantage for location gained by encoding the spatial configuration of the three objects on the screen.

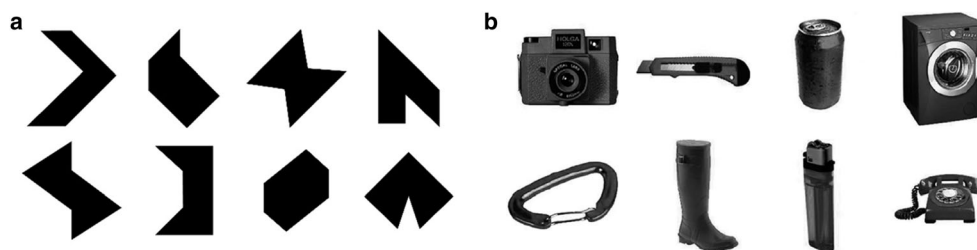
## Method

**Participants** Thirty-two first-year psychology students (24 female) from the University of Sydney participated in the experiment in exchange for course credit. All participants had normal or corrected-to-normal visual acuity and colour vision.

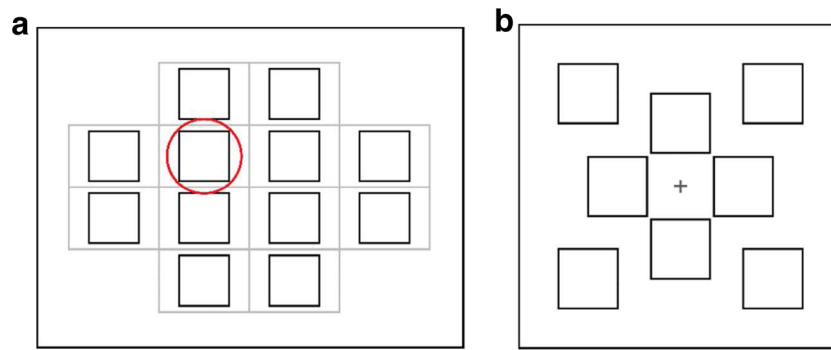
**Apparatus** The experiment was programmed and run in PsychoPy (Peirce, 2007). Participants viewed the display on a 24-inch LCD monitor with a resolution of  $1440 \times 1080$ , set to refresh at 85 Hz. The display measured  $32^\circ \times 24^\circ$  of visual angle at a viewing distance of 70 cm.

**Stimuli** For the abstract stimuli, eight shapes were chosen from work by Parra, Abrahams, Logie, and Sala (2009), as they were demonstrated to be easily visually discriminable but difficult to name (see Fig. 1a). Eight real-world objects were selected from the stimulus set used by Brady, Konkle, Gill, Oliva, and Alvarez (2013) and retrieved from <http://bradylab.ucsd.edu/stimuli/ColorRotationStimuli.zip> (see Fig. 1b). These objects could be rendered in a variety of plausible colours using the MATLAB code accompanying the stimulus set. For this experiment, we generated 12 different coloured variants of each of the real objects and abstract shapes, rotating each item through HSL hue space, and leaving saturation and lightness unchanged. All stimuli measured approximately  $3.7^\circ \times 3.7^\circ$  of visual angle, and were presented in one of 12 locations within the display. These locations formed an invisible  $4 \times 4$  grid, with a minimum of  $4.7^\circ$  of visual angle between the centres of the cells, although the four corner locations were never used (see Fig. 2a).

**Procedure** Each trial began with presentation of a fixation cross for 700 ms, followed by either three abstract shapes or three real-world objects presented in random colours and locations (see Fig. 3). Each stimulus was presented for 300 ms



**Fig. 1** Stimuli used in Experiment 1. **a** Abstract shape stimuli from the set used by Parra et al. (2009). **b** Concrete object stimuli from the set used by Brady et al. (2013)



**Fig. 2** Locations used in the study. **a** Grid of possible stimulus locations used in Experiment 1. Circle indicates size of region considered correct. **b** Grid used for stimulus presentation in Experiment 2

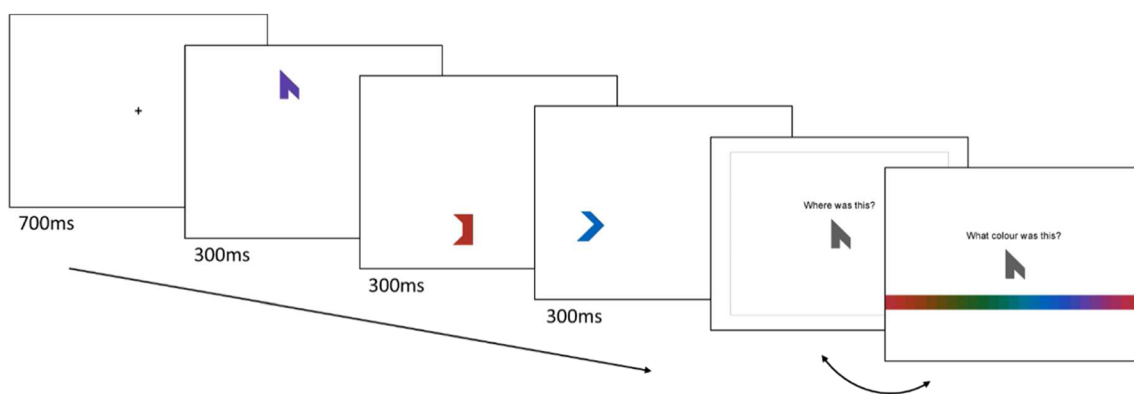
with a 100 ms interstimulus interval. After presentation, participants were prompted with one item in greyscale and asked to report its colour and/or location. For location report, participants could click anywhere on the screen to indicate the rough location of the target; a selection within 3° of visual angle from the centre of the target location was considered correct. For colour report, a bar appeared along the bottom of the screen with each of the 12 possible target colours as well as 13 blended ones to create a more continuous input in line with location responses. To limit the importance of fine-grained colour discrimination, the selection of any of five hues around the actual target was considered correct, and the most similar hues never appeared together in any given trial. This means that the probability of getting the colour correct by chance was 1/5 (0.2), and the probability of getting location correct by chance was 1/12 (0.06).<sup>1</sup> Participants did not receive feedback on their responses.

Participants completed one 10-trial practice block, followed by five experimental blocks. Between each block, participants were instructed which features to attend to. The first

block was always 80 trials using randomly intermixed real-world and abstract stimuli, and participants were asked to report both the colour and location of the probed item, with the order of responses randomised from trial to trial. Following this, they completed four single-feature blocks, each consisting of 30 trials asking for either colour or location and using either abstract or real-world stimuli. The order of these four blocks was counterbalanced across participants.

**Results**

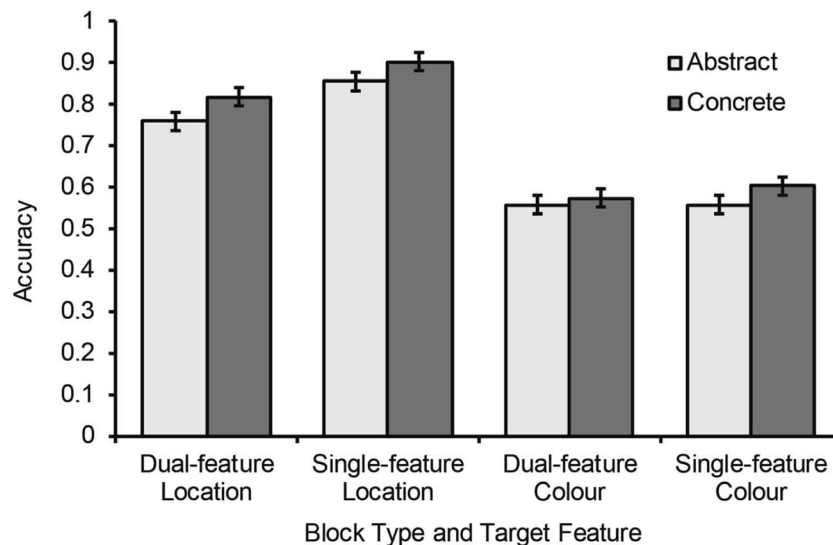
**Dual-feature block** A three-way repeated-measures ANOVA using target feature (colour vs location), stimulus type (real object vs abstract shapes), and response order (colour or location queried first) showed that average accuracy was significantly higher for location ( $M = .789, SD = .100$ ) than for colour reports ( $M = .565, SD = .109$ ),  $F(1, 31) = 300.652, p < .001$ . Overall accuracy was significantly higher for object stimuli ( $M = .696, SD = .098$ ) than for abstract shapes ( $M = .659, SD = .114$ ),  $F(1, 31) = 6.542, p = .016$ ; and higher when



**Fig. 3** Trial structure in Experiment 1

<sup>1</sup> Originally, the number of possible colour and locations were matched, but the difficulty of the colour judgement had to be reduced due to performance being at floor. This discrepancy in chance levels works against the asymmetry in location versus colour accuracy (see Results).

participants were asked for location first ( $M = .712, SD = .107$ ) relative to colour first ( $M = .642, SD = .108$ ),  $F(1, 31) = 20.293, p < .001$ . There were no significant two-way or three-way interactions between these factors, all  $F(1, 31) < 1, p > .05$ . These data are shown in Fig. 4, plotted alongside



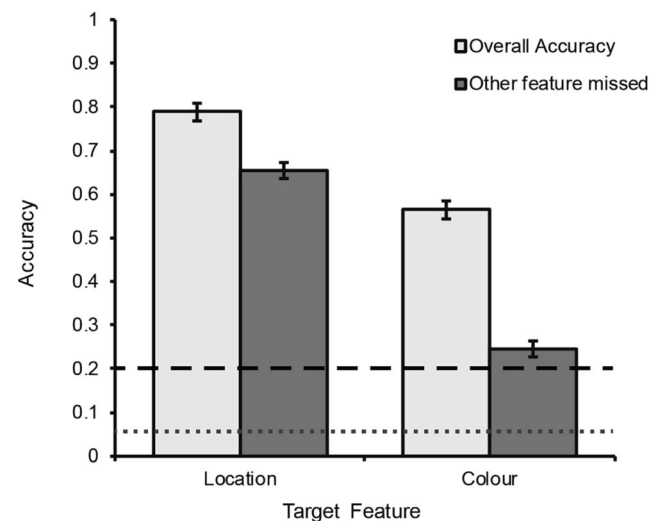
**Fig. 4** Feature report accuracy as a function of block type and stimulus type (concrete, familiar objects, or abstract shapes). Error bars represent within-subjects standard error means (*SEM*) from the three-factor model, as per Loftus and Masson (1994)

data from the single-feature blocks. Given the absence of interactions, data were collapsed across stimulus type and response order for the remaining dual-feature block analyses.

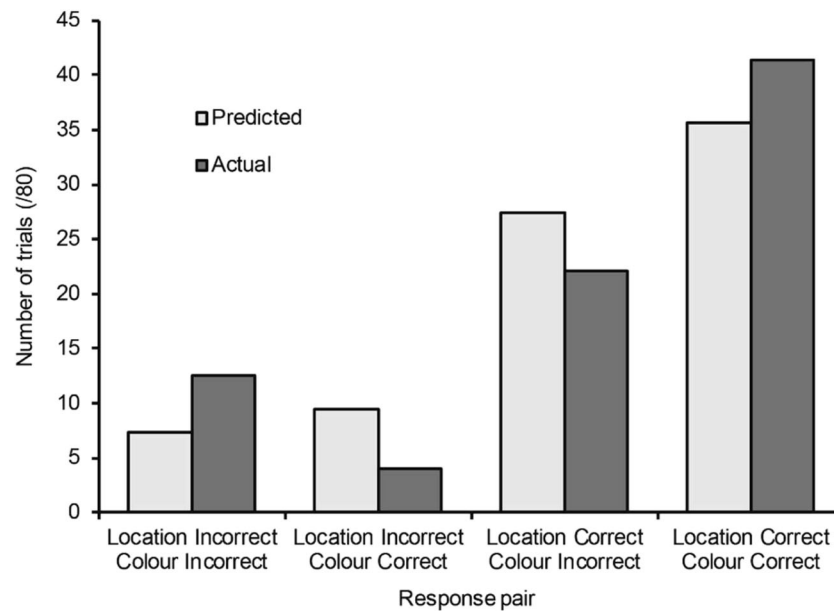
To establish any contingency between colour and location, we next looked at accuracy for each feature on trials in which the other feature was reported incorrectly (i.e., location accuracy on location correct/colour incorrect trials and colour accuracy on colour correct/location incorrect trials) and compared these accuracy rates to chance performance via one-sample *t* tests; here and elsewhere, significance level was Bonferroni-adjusted for the number of comparisons (in this case two, yielding an  $\alpha$  level of .025). On trials where colour was reported incorrectly, location accuracy ( $M = .653$ ,  $SD = .120$ ) remained well above chance level (chance = .06),  $t(31) = 27.91$ ,  $p < .001$ . In contrast, on trials where location was reported incorrectly, colour accuracy ( $M = .245$ ,  $SD = .117$ ) was not significantly different to chance (chance = .20),  $t(31) = 2.18$ ,  $p = .037$  (see Fig. 5), suggesting that participants could only report the colour of the item when they knew its location.

To further test this contingency, we modelled a distribution of types of responses based on the assumption that the two features are coded independently and compared this with the actual distribution of response pairs (see Fig. 6). There were four possible response combinations: both features reported correctly, only location correct, only colour correct, or both features reported incorrectly. For the predicted distribution, overall accuracy for each feature was used to estimate the proportion of trials that would fall into each of the four response categories if the two responses were independent. That is, if location was reported correctly 80% of the time and colour 50% of the time, if the two features were independent, we would expect participants to get both correct on 40% of trials). We used a chi-square test for goodness of fit to compare

the predicted distribution of response pairs with the actual distribution and found a significant deviation from the predicted values,  $\chi^2(3) = 9.90$ ,  $p = .019$ . As illustrated in Fig. 6, participants reported both features correctly or both features incorrectly more often than predicted under the independence hypothesis, and they reported only one feature correctly less often than predicted. This suggests that participants were actually binding the features of the items. However, importantly, the distribution of actual responses clearly shows that colour reports are contingent on getting the location correctly: There were almost no trials in which location was incorrect but colour was correct (4/80), supporting the conclusion that colour report is at chance when location is not known.



**Fig. 5** Feature accuracy overall relative to accuracy when the other feature was reported incorrectly. Dashed line represents chance accuracy for colour feature; dotted line represents chance for location. Error bars represent standard error of the mean difference



**Fig. 6** Comparison of actual distribution of response pairs in the dual-feature block and predicted distribution based on feature report accuracy

**Single-feature blocks** Accuracy data for the single-feature blocks are shown in Fig. 4. A two-way repeated-measures ANOVA with target feature and stimulus type as factors revealed a main effect for target feature,  $F(1, 31) = 389.624$ ,  $p < .001$ , with accuracy being significantly higher for location ( $M = .879$ ,  $SD = .089$ ) than for colour reports ( $M = .581$ ,  $SD = .107$ ). Overall, accuracy was significantly higher for concrete object stimuli ( $M = .753$ ,  $SD = .089$ ) than for abstract shapes ( $M = .707$ ,  $SD = .103$ ),  $F(1, 31) = 12.587$ ,  $p = .001$ . These factors did not interact,  $F(1, 31) < 1$ ,  $p > .05$ .

**Comparison of dual-feature and single-feature blocks** To evaluate whether there were any costs in performance when reporting two features compared with one, we ran an ANOVA with the factors of block type (dual feature vs single feature), target feature (colour vs location), and stimulus type (objects vs abstract shapes). This yielded a main effect for block type,  $F(1, 31) = 20.500$ ,  $p < .001$ , with accuracy in the single-feature blocks ( $M = .730$ ,  $SD = .089$ ) being significantly higher than in the dual-feature blocks ( $M = .677$ ,  $SD = .098$ ). As expected from previous analyses, accuracy was significantly higher for location ( $M = .834$ ,  $SD = .087$ ) than for colour reports ( $M = .573$ ,  $SD = .098$ ),  $F(1, 31) = 611.929$ ,  $p < .001$ , and for real objects ( $M = .724$ ,  $SD = .087$ ) than for abstract shapes ( $M = .683$ ,  $SD = .098$ ),  $F(1, 31) = 14.648$ ,  $p = .001$ .

Importantly, there was a significant interaction between block type and the target feature,  $F(1, 31) = 15.666$ ,  $p < .001$ , with a greater accuracy advantage on single-feature blocks for location than for colour reports (see Fig. 4). Paired-sample  $t$  tests with means collapsed across stimulus type revealed that the single-feature advantage was significant

for location reports,  $t(31) = 6.411$ ,  $p < .001$ , but the single-feature accuracy was equivalent to accuracy in the dual-feature condition in the case of colour reports,  $t(31) = .904$ ,  $p > .05$ . This suggests that there is no additional processing cost associated with encoding location when already reporting the colour. The other two-way and three-way interactions between these factors were not significant.

## Discussion

The results of Experiment 1 reveal a clear contingency between reports of colour and reports of location, consistent with the predictions of location-based binding accounts. In dual-feature trials where location was reported incorrectly, colour accuracy was at chance levels. However, the inverse was not true; knowledge of location did not appear to be contingent on colour, with location reports remaining substantially above chance when colour was incorrectly reported. Our results do indicate that it was easier to report location than colour, and as such, it is conceivable that this asymmetry could be due to a greater likelihood that colour be reported incorrectly on any trial where the participants were less attentive or alert. However, the goodness-of-fit analysis suggests that relative task difficulty cannot account for our results, and that reports of the two features are unlikely to be independent. Importantly, the decrease in colour accuracy was not only much larger than for location, but it was down to chance levels, suggesting an inability to process colour in isolation, rather than a mere bias towards location report. Additionally, accuracy of location reports was higher in blocks where participants did not need to report colour, reflecting a processing cost of the colour feature, but there was no equivalent

advantage when only colour had to be reported, and any bias towards reporting the easier feature could have boosted performance. This suggests either that participants still needed to attend to location when they reported colour only, as predicted by indexing theories, or that there is no processing cost for attending to location, possibly due to automatic registration from eye-movements. This latter possibility was tested in Experiment 2.

Participants accurately reported both features or missed both features of a target more frequently than predicted by accuracy for each feature on its own (see Fig. 6), suggesting a tendency to bind the features together. While this is superficially consistent with strong accounts of binding, the relatively high proportion of location-correct responses on trials where one feature was missed points instead to asymmetric binding processes. Our findings are thus more consistent with an asymmetric model in which participants are able to report the dominant feature (location) on its own, but are unable to access the dependent feature on trials where location was missed (Rangelov & Zeki, 2014).

Findings from the order in which features were queried provide further support for the structure proposed by location-based theories. Accuracy for both colour and location responses was significantly higher when location was queried first, relative to when colour was queried first. Object indexing theories claim that objects are identified by their location, and by drawing attention to the location of targets first, participants appear to have better access to all of the target features.

Although accuracy across all conditions was higher for familiar-object stimuli, the lack of interactions with other factors suggests this advantage is not the result of distinct binding processes. This finding is inconsistent with the idea that familiar objects without overriding colour associations would be less reliant on location, despite not using hardwired conjunctions (Hommel & Colzato, 2009; VanRullen, 2009). Rather, there appeared to be a general processing advantage which may be due to ease of phonological rehearsal when naming the real-world objects. Alternatively, the familiar objects may have simply been more memorable. In any case, they did not appear to engage location in binding differently to the abstract stimuli.

## Experiment 2

The aim of Experiment 2 was to determine whether the apparent importance of location in binding is due to feedback from eye movements. A similar setup to Experiment 1 was used, but participants were required to maintain fixation during half of the trials, and an eye tracker was used to confirm that they did not break fixation. The visual display was condensed to  $10^\circ$  of visual angle to fit comfortably within the radius of intact colour perception (conservative estimate of  $20^\circ$  from

Hansen, Pracejus, & Gegenfurtner, 2009) around the centre of the display. To offset the difficulty increase from having to maintain fixation, we only used eight possible colours and locations, rather than the 12 from Experiment 1, and stimuli were presented for slightly longer. This experiment only used the abstract shape stimuli.

## Method

**Participants** Twenty (16 female) new participants were recruited from the same pool as Experiment 1.

**Apparatus** Experiment 2 was conducted on a 17-inch CRT monitor, with a refresh of 85 Hz and a resolution of  $1024 \times 768$ . Eye movements were recorded using an SR Research Eyelink 1000 controlled with functions called from PsychoPy code written by Chen and Fajou (2015). A desk-mounted head brace restricted movement and maintained a viewing distance of 60 cm and experimental display of  $10^\circ \times 10^\circ$  of visual angle.

**Stimuli** The same eight abstract shape stimuli from Experiment 1 were used, rendered in eight equally distinct hues using the MATLAB code provided by Brady et al. (2013). All stimuli measured  $2^\circ \times 2^\circ$  of visual angle and could be presented in any of eight possible locations (see Fig. 2b).

**Procedure** The procedure was similar to Experiment 1. After calibration of the eye tracker, participants completed a 10-trial practice block asking for both colour and location, followed by six experimental blocks. In half of the blocks participants were free to move their eyes naturally, but in the other half they were required to maintain fixation. The order of these was counterbalanced across participants. In both the restricted and nonrestricted (free) eye-movement conditions, a dual-feature block was administered first, followed by the two single-feature blocks in a counterbalanced order. For the eye-movement-restricted block, trials were discarded if participants broke fixation ( $1^\circ$  of visual angle around the centre of the screen) or blinked during stimulus presentation. The free eye-movement conditions included a 60-trial dual-feature block and two 30-trial single-feature blocks. To compensate for possible lost trials, the restricted eye-movement conditions had 70 trials in the dual-feature block and 35 trials in each of the single-feature blocks.

During each trial, a fixation cross was presented for 1,600 ms before the stimuli sequence, to allow participants time to settle their eyes, and the cross persisted throughout the trial to make it easier to maintain fixation. Three differently coloured stimuli were sequentially presented for 320 ms with a 130-ms interstimulus interval. After the sequence, participants were prompted for colour and/or location as in



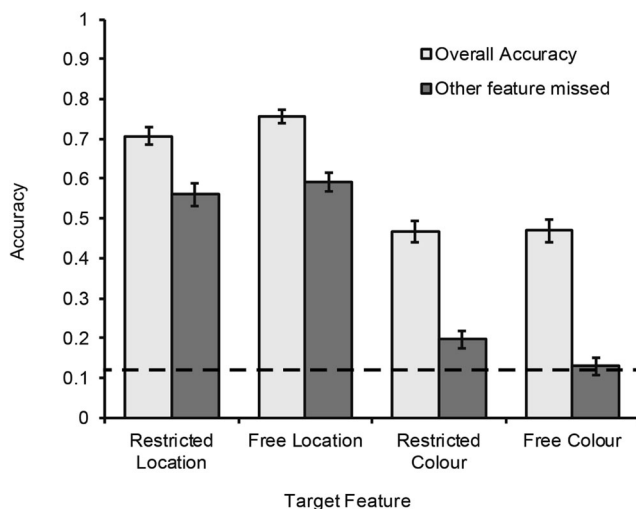
Experiment 1, and were free to move their eyes during response selection.

## Results

For the restricted eye-movement blocks, participants successfully maintained fixation on 65% of trials in the dual-feature block, and 74% of trials in both of the single-feature blocks. On trials in which fixation was broken, participants made an average of 5.66 ( $SD = 2.85$ ) eye movements, tending to look at items as they appeared, before returning to fixation.

**Dual-feature blocks** A three-way repeated-measures ANOVA with the factors of target feature (colour vs location), order (colour or location queried first) and eye movement (restricted or free) revealed a main effect of target feature,  $F(1, 19) = 214.557$ ,  $p < .001$ , with higher accuracy for location ( $M = .732$ ,  $SD = .053$ ) than for colour reports ( $M = .469$ ,  $SD = .113$ ). Accuracy was higher when location was queried first ( $M = .638$ ,  $SD = .088$ ) than when colour was queried first ( $M = .563$ ,  $SD = .084$ ),  $F(1, 19) = 14.087$ ,  $p = .001$ . Both of these results replicate the findings of Experiment 1. There was no overall effect of eye movements,  $F(1, 19) = .721$ ,  $p > .05$ . However, there was a significant interaction between eye movement and the target feature,  $F(1, 19) = 5.091$ ,  $p = .036$ , with the advantage for location accuracy relative to colour accuracy being greater when participants could move their eyes freely. There were no other significant interactions, so the data were collapsed across response order for the subsequent analyses.

We next tested for contingency between the features by looking at accuracy for one of the features when the other feature was reported incorrectly in the same manner as in Experiment 1 (see Fig. 7). Four one-sample  $t$  tests were



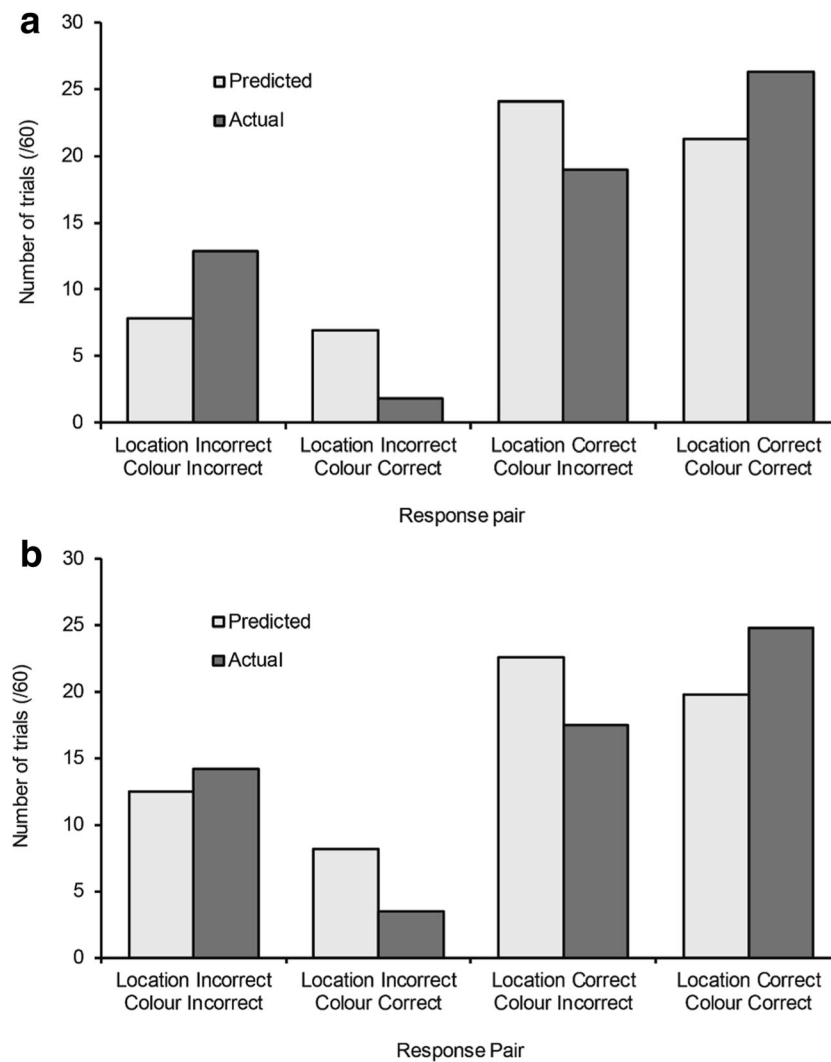
**Fig. 7** Feature accuracy overall relative to accuracy when the other feature was reported incorrectly in both free eye-movement and restricted eye-movement conditions for the dual-feature blocks. Dashed line represents chance accuracy. Error bars represent standard error of the mean difference

performed to check whether accuracy for each feature was significantly different from chance (Bonferroni-corrected  $\alpha$  level of .0125). On trials where colour was reported incorrectly, location accuracy was well above chance (chance = .125) in both the free condition ( $M = .591$ ,  $SD = .107$ ),  $t(19) = 19.52$ ,  $p < .001$ , and in the restricted condition ( $M = .560$ ,  $SD = .125$ ),  $t(19) = 15.55$ ,  $p < .001$ . For trials where location was reported incorrectly, colour accuracy was not significantly different to chance (chance = .125) in the free condition ( $M = .130$ ,  $SD = .099$ ),  $t(19) = .21$ ,  $p > .05$ , and dropped to near chance in the restricted condition ( $M = .197$ ,  $SD = .101$ ), though the difference was still significant after Bonferroni correction,  $t(19) = 3.19$ ,  $p = .005$ .

As in Experiment 1, we compared the distribution of actual responses to the distribution predicted on the basis of independent processing of features (see Fig. 8). Chi-square tests for goodness of fit showed significant deviation from the predicted values in both the nonrestricted,  $\chi^2(3) = 8.93$ ,  $p = .030$ , and restricted eye-movement conditions,  $\chi^2(3) = 12.24$ ,  $p = .007$ . The pattern of response distributions is very similar to that found in Experiment 1, and does not change substantially according to the eye-movement condition. It is clear, once again, that participants gave very few correct colour answers when location was reported incorrectly.

**Single-feature blocks** A two-way repeated-measures ANOVA with eye-movement restriction and target feature as factors revealed a significant main effect of eye movement,  $F(1, 19) = 6.090$ ,  $p = .023$ ; accuracy was modestly, though significantly, higher when eye movements were permitted ( $M = .665$ ,  $SD = .093$ ) relative to when participants were required to maintain fixation ( $M = .611$ ,  $SD = .107$ ). Average accuracy was significantly higher for location ( $M = .795$ ,  $SD = .076$ ) than for colour report ( $M = .481$ ,  $SD = .136$ ),  $F(1, 19) = 109.992$ ,  $p < .001$ . There was no interaction between eye-movement restriction and the target feature,  $F(1, 19) < 1$ ,  $p > .05$ .

**Comparison of dual-feature and single-feature blocks** A final ANOVA (see Fig. 9) using block type (single vs dual feature), eye-movement restriction, and target feature as factors revealed that accuracy was significantly higher when eye movements were permitted ( $M = .639$ ,  $SD = .078$ ) relative to when participants were required to maintain fixation ( $M = .599$ ,  $SD = .089$ ),  $F(1, 19) = 5.001$ ,  $p = .038$ . Accuracy was also significantly higher in the single-feature blocks ( $M = .638$ ,  $SD = .088$ ) relative to the dual-feature blocks ( $M = .600$ ,  $SD = .074$ ),  $F(1, 19) = 6.793$ ,  $p = .017$ . Finally, accuracy was significantly higher for location ( $M = .763$ ,  $SD = .056$ ) than for colour reports ( $M = .475$ ,  $SD = .108$ ),  $F(1, 19) = 216.035$ ,  $p < .001$ . There were no significant two-way or three-way interactions between these factors, all  $F$ s  $< 1$ ,  $p > .05$ . To check for replication of the results from Experiment 1, two additional paired-sample  $t$  tests were conducted with means collapsed across



**Fig. 8** Comparison of actual and predicted distribution of response pairs in the dual-feature block with (a) free eye movement and (b) restricted eye movement

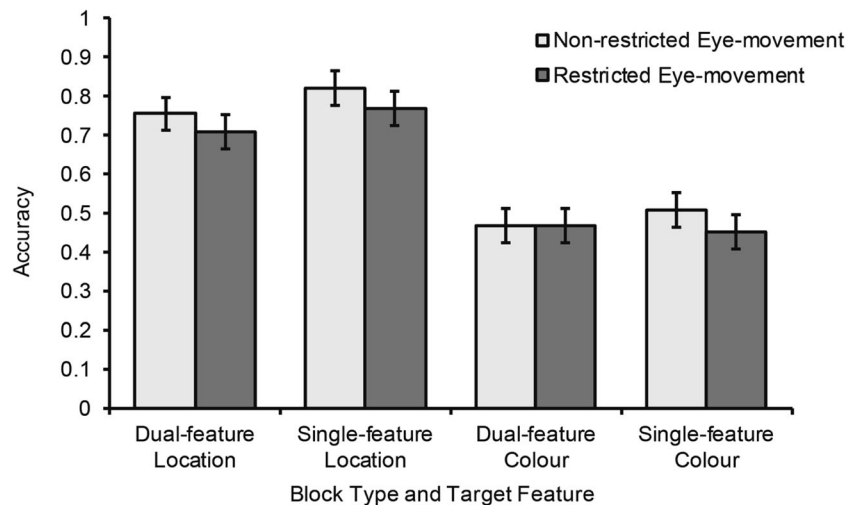
eye-tracking conditions, comparing accuracy for each feature under single and dual-feature report. As in Experiment 1, these tests revealed a significant single-feature advantage for location,  $t(19) = -4.085$ ,  $p < .001$ , but no equivalent advantage for colour,  $t(19) = .523$ ,  $p > .05$ .

## Discussion

Feedback from eye movements does not appear to drive the general location advantage found in Experiment 1 and replicated in Experiment 2. Although location accuracy was significantly more impaired than colour accuracy by the restriction of eye movements in the dual-feature block, the effect was modest, and this relationship was not found to be significant in the single-feature blocks, or in the interaction between the single-feature and dual-feature blocks. A concern might be that this is because participants were not in fact moving their eyes during the free-viewing condition. This was not the case.

Whenever participants failed to maintain fixation, they tended to move their eyes towards the objects as they appeared, making on average 5–6 eye movements per trial. There were also overall differences in accuracy between the free-viewing and restricted conditions, consistent with differences in eye-movement behaviour. However, the location contingency in binding does not appear to be driven by eye movements. As in Experiment 1, colour accuracy fell to near chance levels on trials in which location was reported incorrectly, regardless of whether or not participants could move their eyes. This suggests that the apparent contingency between colour and location is not because location is automatically registered from eye movements when attending to colour.

The accuracy advantage for both colour and location when location was queried first was demonstrated again and was found to be insensitive to whether eye movements were permitted during study. If this advantage had been absent when eye movements were restricted, we could have concluded that



**Fig. 9** Feature report accuracy as a function of block type and eye movement. Error bars represent within-subjects *SEM* from the three-factor model, as per Loftus and Masson (1994)

it was driven by a match between encoding and retrieval, in line with the encoding specificity principle (Tulving & Thomson, 1973). That is, querying location first could cause the eyes to return to the position they were in when the stimulus was originally studied, and this match would lead to better feature report accuracy. However, this was not the case. This suggests that the effect is due to better access to individual object features via location, consistent with location indexing theories. As in Experiment 1, our data seem most consistent with asymmetric access to the probed features, with colour reports being contingent upon successful location report.

## General discussion

The current study aimed to determine whether direct binding occurs between visual features, or whether feature conjunctions are necessarily mediated by location. The results of both experiments suggest that binding is reliant on location information, as predicted by FIT and object indexing theories.

Experiment 1 contradicted the claims of Allen et al. (2015)—both that direct binding is generally possible and the caveat that it may only be used according to certain task demands. It may be worth noting that in Allen et al.’s study, location targets were cued by an arrow pointing to their place in the original four-item array, but on collocated trials, the last thing occupying the indicated space was the interfering suffix. Therefore, an alternative explanation for their finding that collocated visual suffixes produced differential interference only when also cued by location is that location cues increased the salience of the suffixes, and consequently their interference, rather than reflecting a change in binding strategy. Indeed, across our experiments, when location was reported

incorrectly, colour accuracy was close to chance levels, indicating that participants were very unlikely to have bound the colour information directly to the shape. The inverse was not true; participants frequently reported location correctly when colour was reported incorrectly, suggesting that colour encoding is contingent upon location, but not vice versa. This contingency persisted even as task demands were changed; a switch to direct binding processes in the single-feature blocks, where participants no longer needed to report location, should have led to an increase in colour accuracy, but we saw no evidence for this. This apparently location-dependent colour-shape binding is consistent with the results of Schneegans and Bays (2017), who described a similar reliance on location when reporting a stimulus feature (colour) when cued with the other feature (orientation). Experiment 2 further ruled out the possibility that direct binding might be used when automatically captured location information from eye movements is unavailable. There was no indication that participants were ever able to ignore location in binding, consistent with the automatic registration described by Chen and Wyble (2015), and in line with the predictions of location-based theories. However, as our design relied on short-term memory for the targets, it is not possible to rule out postperceptual binding in working memory, as proposed by Pertzov and Husain (2014).

The results of this study do not support the proposal that binding for familiar objects is less reliant on location information than binding the features of unfamiliar objects. Across all conditions of Experiment 1, accuracy for the colour and location of familiar object stimuli was higher than accuracy for unfamiliar shape stimuli. However, there was no significant interaction between stimulus type and the target features in any condition, indicating that access to colour for the familiar stimuli used in this experiment relied on location to the same extent as for the unfamiliar abstract shapes. Different binding

processes might be engaged for familiar objects with consistent real-world colour associations, unlike the arbitrary associations used here, but simply using familiar objects does not appear to reduce location reliance when binding is required.

The second experiment aimed to test whether the processing advantage for location was driven by feedback from eye movements rather than being due to the manner in which objects are represented. We found little evidence to suggest such a role for eye movements. The substantial drop in colour accuracy when location was reported incorrectly persisted when eye movements were restricted, and once again there was no equivalent reduction in location report accuracy when colour was missed. This suggests that the apparent location-colour contingency arises at the cognitive level; there is no motor component making location information more available. Having found no evidence for location feedback from eye movements also suggests that the lack of any apparent processing cost for attending to location in addition to colour is not because location comes for free, due to the motor signals from eye position. Rather, it appears that visual feature binding is actually reliant on location in the manner predicted by location indexing theories. This contingency also provides an explanation as to why difficulty reporting each feature could not be matched, despite repeatedly making colour report easier.

Evidence against a feedback role for eye movements suggests that the abnormal binding performance in Balint's syndrome is rooted in spatial processing deficits, rather than a coincident feature arising from atypical eye movements. Had the present study found that location information is derived from eye movements, difficulties localising the parts of objects (Robertson & Treisman, 2006) and abnormal visual search performance (Coslett & Lie, 2008) could have been the indirect result of oculomotor apraxia. However, location information was not found to be derived from eye movements, indicating that the current separation of these symptoms is valid.

Taken together, the results from these two experiments are strong support for object indexing theories. Location information appears to play a crucial role in binding, mediating access to other features and driving the representation of objects. We found no evidence that direct binding between colour and shape takes place, regardless of task demands. This provides a good explanation for the seemingly automatic registration of location; participants cannot ignore location because they need it in order to encode other visual features. Finally, these findings cannot be attributed to feedback from eye movements; rather, they appear to be indicative of the structure of object representations used in perception or memory.

**Acknowledgements** This research was supported by a Future Fellowship (FT0992123) from the Australian Research Council awarded to I.M. Harris. The authors declare no conflicts of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Allen, R. J., Castellà, J., Ueno, T., Hitch, G. J., & Baddeley, A. D. (2015). What does visual suffix interference tell us about spatial location in working memory? *Memory & Cognition*, *43*(1), 133–142. <https://doi.org/10.3758/s13421-014-0448-4>
- Arnold, D. H. (2005). Perceptual pairing of colour and motion. *Vision Research*, *45*(24), 3015–3026. <https://doi.org/10.1016/j.visres.2005.06.031>
- Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631. <https://doi.org/10.1016/j.neuropsychologia.2010.12.023>
- Bosbach, S., Kornblum, C., Schröder, R., & Wagner, M. (2003). Executive and visuospatial deficits in patients with chronic progressive external ophthalmoplegia and Kearns–Sayre syndrome. *Brain*, *126*(5), 1231–1240. <https://doi.org/10.1093/brain/awg101>
- Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*(6), 981–990. <https://doi.org/10.1177/0956797612465439>
- Brockmole, J. R., & Franconeri, S. L. (2009). Introduction. *Visual Cognition*, *17*(1/2), 1–9. <https://doi.org/10.1080/1350628080233211>
- Bruce, S., & Muhammad, Z. (2009). The development of object permanence in children with intellectual disability, physical disability, autism, and blindness. *International Journal of Disability, Development and Education*, *56*(3), 229–246. <https://doi.org/10.1080/10349120903102213>
- Chen, W.-Y., & Fajou, C. (2015). EyeLinkForPsychopyInSUPA [Computer software]. Retrieved from <https://github.com/alexholcombe/MOTcircular/blob/master/EyelinkeYetrackerForPsychopySUPA3.py>
- Chen, H., & Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, *107*, 76–85. <https://doi.org/10.1016/j.visres.2014.11.010>
- Clifford, C. W. G., Holcombe, A. O., & Pearson, J. (2004). Rapid global form binding with loss of associated colors. *Journal of Vision*, *4*(12), 1090. <https://doi.org/10.1167/4.12.8>
- Coslett, H. B., & Lie, G. (2008). Simultanagnosia: When a rose is not red. *Journal of Cognitive Neuroscience*, *20*(1), 36–48. <https://doi.org/10.1162/jocn.2008.20002>
- Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, *269*(5225), 853–855.
- Golomb, J. D., Kupitz, C. N., & Thiemann, C. T. (2014). The influence of object location on identity: A “spatial congruency bias”. *Journal of Experimental Psychology: General*, *143*(6), 2262–2278. <https://doi.org/10.1037/xge0000017>
- Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision*, *9*(4), 26–26.
- Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, *13*(5), 216–221. <https://doi.org/10.1016/j.tics.2009.02.005>

- Holcombe, A. O., & Clifford, C. W. G. (2012). Failures to bind spatially coincident features: Comment on Di Lollo. *Trends in Cognitive Sciences*, 16(8), 402–402. <https://doi.org/10.1016/j.tics.2012.06.011>
- Hommel, B., & Colzato, L. S. (2009). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition*, 17(1/2), 120–140. <https://doi.org/10.1080/13506280802349787>
- Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York, NY: Academic Press.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11), 1762–1776. <https://doi.org/10.1016/j.visres.2005.10.002>
- Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing ‘what’ and ‘where’ systems. *Trends in Cognitive Sciences*, 2(1), 10–18. [https://doi.org/10.1016/S1364-6613\(97\)01113-3](https://doi.org/10.1016/S1364-6613(97)01113-3)
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490.
- Meegan, D. V., & Honsberger, M. J. M. (2005). Spatial information is processed even when it is task-irrelevant: Implications for neuroimaging task design. *NeuroImage*, 25(4), 1043–1055. <https://doi.org/10.1016/j.neuroimage.2004.12.061>
- Parra, M. A., Abrahams, S., Logie, R. H., & Sala, S. D. (2009). Age and binding within-dimension features in visual short-term memory. *Neuroscience Letters*, 449(1), 1–5. <https://doi.org/10.1016/j.neulet.2008.10.069>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods*, 162(1), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Pertzov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception, & Psychophysics*, 76(7), 1914–1924.
- Rajic, J., & Wilson, D. E. (2014). Asymmetrical access to color and location in visual working memory. *Attention, Perception, & Psychophysics*, 76(7), 1902–1913. <https://doi.org/10.3758/s13414-014-0723-2>
- Rangelov, D., & Zeki, S. (2014). Non-binding relationship between visual features. *Frontiers in Human Neuroscience*, 8, 749. <https://doi.org/10.3389/fnhum.2014.00749>
- Robertson, L. C., & Treisman, A. (2006). Attending to space within and between objects: Implications from a patient with Balint’s syndrome. *Cognitive Neuropsychology*, 23(3), 448–462. <https://doi.org/10.1080/02643290500180324>
- Rodriguez, V., Valdes-Sosa, M., & Freiwald, W. (2002). Dividing attention between form and motion during transparent surface perception. *Cognitive Brain Research*, 13(2), 187–193. [https://doi.org/10.1016/S0926-6410\(01\)00111-2](https://doi.org/10.1016/S0926-6410(01)00111-2)
- Rogers, B. (2009). Motion parallax as an independent cue for depth perception: A retrospective. *Perception*, 38(6), 907. <https://doi.org/10.1068/p080125>
- Ryan, J. D., & Villate, C. (2009). Building visual representations: The binding of relative spatial relations across time. *Visual Cognition*, 17(1/2), 254. <https://doi.org/10.1080/13506280802336362>
- Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience*, 37(14), 3913–3925. <https://doi.org/10.1523/JNEUROSCI.3493-16.2017>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, 34(8), 1704–1719.
- Troscianko, T., Montagnon, R., Clerc, J. L., Malbert, E., & Chanteau, P. (1991). The role of colour as a monocular depth cue. *Vision Research*, 31(11), 1923–1929. [https://doi.org/10.1016/0042-6989\(91\)90187-A](https://doi.org/10.1016/0042-6989(91)90187-A)
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373. <https://doi.org/10.1037/h0020071>
- Valdes-Sosa, M., Cobo, A., & Pinilla, T. (1998). Transparent motion and object-based attention. *Cognition*, 66(2), B13–B23. [https://doi.org/10.1016/S0010-0277\(98\)00012-2](https://doi.org/10.1016/S0010-0277(98)00012-2)
- VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*, 17(1/2), 103–119. <https://doi.org/10.1080/13506280802196451>
- Walsh, V., Ashbridge, E., & Cowey, A. (1998). Cortical plasticity in perceptual learning demonstrated by transcranial magnetic stimulation. *Neuropsychologia*, 36(1), 45–49. [https://doi.org/10.1016/S0028-3932\(97\)00111-5](https://doi.org/10.1016/S0028-3932(97)00111-5)
- Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158), 679–684. <https://doi.org/10.1038/331679a0>